DOI:
10.1080/13854046.2015.1107137

Document Version
Peer reviewed version

The Minimum Clinically Important Difference in the Repeatable Battery for the Assessment of Neuropsychological Status

Rachel Phillips[1*], Qi Gao[2, 3], Simon Lowes Collinson[2, 4], Audrey Ling[2, 3], Lei Feng[2, 3], Yin Bun Cheung [5, 6], Tze-Pin Ng[2, 3].

1. Biostatistics Department, Singapore Clinical Research Institute, Singapore 138669.
2. Gerontological Research Programme, Faculty of Medicine, National University of Singapore, Singapore.
3. Department of Psychological Medicine, National University of Singapore, Singapore 119228.
4. Department of Psychology, National University of Singapore, Singapore 117570.
5. Centre for Quantitative Medicine, Duke-National University of Singapore Graduate Medical School, Singapore 169856.
6. Department of International Health, University of Tampere, Finland 33014.

*Corresponding author

Contact details of authors

RP: Telephone: (65) 6508 6754
Email: rachel.phillips@scri.edu.sg
QG: Telephone: (65) 6772 2286
Email: pcmgqi@nus.edu.sg
SLC: Telephone: (65) 6516 8117
Email: psycsl@nus.edu.sg
AL: Telephone: (65) 6772 3492
Email: pcmlhla@nus.edu.sg
LF: Telephone: (65) 6772 3491
Email: lei_feng@nuhs.edu.sg
YBC: Telephone: (65) 6576 7379
Email: yinbun.cheung@duke-nus.edu.sg
TPN: Email: (65) 6772 3478
Telephone: tze_pin_ng@nuhs.edu.sg

Abstract word count: 248
Manuscript word count: 4459

This work has been support by the Biomedical Research Council (NTP, 03/1/21/17/214, 08/1/21/19/567). The authors report there are no conflicts of interest.

Abstract

Objective: There is no established minimum clinically important difference (MCID) for the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) index and total scale scores. This study aimed to estimate the MCID for the RBANS index scores and total scale score.

Method: Participants included 1,856 ethnic Chinese, older adults. Distribution and anchor-based methods were used to estimate values for the MCID. Distribution-based estimates were calculated as the standard error of measurement (SEM) and 0.5 standard deviations (SD). For anchor-based estimates we compared RBANS scores between the clinical dementia rating (CDR) scale no dementia and very mild dementia groups and between the clinical assessment of dementia (CAD) cognitively normal and mild cognitive impairment groups using regression models adjusting for demographic characteristics.

Results: Estimates from the CDR anchor were 7.79, 8.63, 10.74, 9.74, 5.61 and 3.77 for the total scale score, language, immediate memory, delayed memory, visuospatial/constructional and the attention index, respectively. Estimates from the distribution-based methods were similar to the estimates based on the CDR, except for the language and attention indexes. Estimates from the clinical assessment of dementia (CAD) anchor were larger.

Conclusions: We estimated the MCID for the total scale score, language, immediate memory, delayed memory, visuospatial/constructional and attention indexes of the RBANS as 8, 9, 10, 10, 6 and 4 points, respectively. These estimates are best suited to discriminate between patient groups, in for example, a clinical trial setting. Further research is needed using longitudinal data to assess their applicability to assess within patient differences.

Keywords: Neuropsychological test, Elderly/Geriatrics/Ageing, Mild cognitive impairment, minimum clinically important difference, Chinese

Introduction

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) was designed to identify and characterize abnormal cognitive decline in older adults (Randolph, 1998). The RBANS has also been shown to be sensitive to detecting impairments in patients with mild cognitive impairment (MCI) (Badenes, Casas, Cejudo & Aguilar, 2008; Mahncke et al., 2006; Kotani et al., 2006). The battery is comprised of 12 subtests that assess the domains of immediate and delayed memory, language, attention and visuospatial/construction (Randolph, 1998). The concise structure of the RBANS and its ease of use make it an attractive tool for older patients compared to lengthier and more difficult neuropsychological assessments (Randolph, Tierney, Mohr & Chase, 1998). These characteristics also make it a useful battery for clinicians evaluating elderly patients with abnormal cognitive decline. However, to date there is no direct clinical interpretation for the numerical scores derived from the RBANS that can be applied objectively and systematically in either a clinical trial or clinical practice setting.

Clinical trials that compare two groups of patients receiving different treatments or interventions conventionally report their results in terms of statistical significance, which is a mathematical way of expressing the likelihood that an observed difference is unlikely to have been caused by chance. However, statistical significance does not necessarily equate to clinical significance, which refers to the practical importance of the observed difference. Clinical significance also plays a role in interpreting results in clinical practice with regards to instruments such as the RBANS, which are based on summary scores across multiple dimensions. Interpreting results from such instruments can lack objectivity if clinicians must rely on personal experience with individual patients and populations (Hermes, Sokoloff, Stroup & Rosenheck, 2012). The minimum clinically important difference (MCID) has been proposed as a more objective way of establishing clinical relevance to changes in standardized instrument scores such as the RBANS and can be used in assessing the effectiveness of a treatment. Guyatt,

Walter and Norman (1985) highlighted the importance of establishing the MCID of an instrument in order to assess the instrument's responsiveness to change. Since then there have been several proposed definitions of the MCID. The most quoted definition comes from Jaeschke, Singer and Norman (1989) who described it as "the smallest difference … which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management." MCID depends on patient populations, therefore there are often a range of estimates for a specific instrument (Revicki, Hays, Cella & Sloan, 2008).

A number of methods have been established over recent years to estimate the MCID. Lydick and Epstein (1993) provided an important taxonomy for these methods: anchor-based methods and distribution-based methods. Anchor-based methods compare the standardized instrument scores to some external criterion, which may be clinician or patient rated such as the Clinical Dementia Rating (CDR) scale (Hughes, Berg, Danziger, Coben & Martin, 19820). Distribution-based methods estimate the MCID based on some measure of variability of the observed scores, such as the standard error of measurement (SEM), standard deviation (SD) or the effect size (Copay, Subach, Glassman, Polly & Schuler, 2007). Copay et al. (2007) and King (2011) have both reviewed and summarized the most popular methods within these categories in use today. However, there is no clear consensus on the best estimation method and current best practice advocates the use of multiple methods to estimate the MCID (Guyatt, Osoba, Wu, Wyrwich & Geoffrey, 2002; Revicki et al., 2008). Revicki et al. (2008) advised to base the MCID, "primarily on relevant patient-based and clinical anchors, with clinical trial experiences to further inform understanding", and that distribution-based methods should be used to "support and help interpret estimates from anchor-based approaches", or when no anchor-based estimates are available. This in turn was recommended by the US FDA in their guidance for industry on patient reported outcomes (Food and Drug Administration, 2009) and

the statement of the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) (Dworkin et al., 2008).

The MCID has already been established for a number of instruments based on summary scores across multiple dimensions similar to the RBANS. For example, Kohn, Sidovar, Kaur, Zhu & Coleman (2014) established a range of MCIDs using distribution based methods for the EuroQol 5-dimension (EQ-5D) health status index in persons with multiple sclerosis; and Y.T. Cheung et al., (2014) used both anchor and distribution based techniques to establish the MCID for the functional assessment of cancer therapy: cognitive function (FACT-Cog) in breast cancer patients. In a study that compared the English original versus the Chinese translation of a quality of life questionnaire, a distribution based approach was used to define a MCID as 0.25 SD (Cheung et al., 2004). Having controlled for potential confounders, two physical well-being item scores were found statistically significantly different between respondents to the English and Chinese versions: one of them had a mean difference of 0.41 SD whereas the other only differed by 0.16 SD. The former was concluded as clinically significantly different between the two language versions but that latter was not.

The Singapore Longitudinal Ageing Study (SLAS- II) is an ongoing population based study of ageing, community dwelling adults in Singapore. The aim of this study was to estimate the MCID of the RBANS for each index score and the total scale score using both anchor and distributive-based techniques. As only cross-sectional data was available for the analysis, MCID estimates were based on between patient differences; estimates based on change over time within patient were not possible.

Methods

Sample

The design and methods of SLAS-II have been described previously (Lim, Collinson, Feng & Ng, 2010). In summary, trained research nurses conducted recruitment via door-to-door visits between March 2008 and October 2013. Consenting participants made an appointment to visit the research center at a later date to receive a battery of neuropsychological and clinical assessments, including the RBANS. Those with a Mini-Mental State Examination (MMSE) score < 26 on the locally validated Singaporean, translated, and modified version (Ng, Niti, Chiam & Kua, 2007) underwent further clinical assessments, including the CDR scale (Hughes et al., 1982), clinical and laboratory tests and a final panel review for diagnosis of Alzheimer's Disease (AD) or other dementias (Nyunt et al., 2013). Feng, Chong, Lim and Ng (2012) have previously shown that 26 is the optimal cut-off on the local MMSE version for detecting early cognitive impairment. The small number of non-Chinese participants (Malays, Indians, and others) recruited into the SLAS-II study rendered inferences regarding ethnicity meaningless, hence they were excluded from this analysis. The study was approved by the National University of Singapore (NUS) Institutional Review Board (IRB).

Measures

The RBANS (Form A) was administered to the participants once in the language–dialect (English, Mandarin, Hokkien, Teochew or Cantonese) according to their dominant or habitual preference, by trained research nurses who were fluent in that particular language–dialect (Collinson, Fang, Lim, Feng & Ng, 2014). Mandarin is the official Chinese language used in Singapore and China. Hokkien, Teochew and Cantonese are dialects commonly used in Singapore and southern China. The translation procedure by a committee of trained multilingual research psychologists has previously been described (Collinson et al., 2014) and the equivalence of these translated test scores to the original English version has been investigated and reported (Phillips et al., 2015). Individual test performance was scored

according to standardized instructions (Randolph, 1998), except for the figure copy and figure recall subtests, which were scored according to the widely accepted, modified criteria, suggested by Duff et al. (2003) to give the 12 raw subtest scores. These 12 raw subtest scores were then standardized across age categories (54 to 59 years old, 60 to 64 years old, 65 to 69 years old, 70 to 74 years old and 75 years and above) using the means and standard deviations produced by Collinson et al. (2014) and then transformed to scores with a mean of 100 and standard deviation of 15. Collinson et al. (2014) also scored the individual test performance as per the standardized instructions proposed by Randolph (1998), with the exception of the figure copy and figure recall subtest which used the modified criteria proposed by Duff et al. (2003). Population specific norms were used as opposed to the standard US norms due to evidenced differences in neuropsychological performance between Asian and Caucasian populations (Boone, Victor, Wen, Razani & Ponto´n, 2007; Fuji, 2010; Hedden et al., 2002; Shan, Chen, Lee & Su, 2008). The five index scores (immediate memory, delayed memory, language, attention and visuospatial/construction) were then calculated by summing the transformed scores of the subtests that contribute to that specific index and taking the mean. The total scale score was derived by taking the mean of the sum of the five index scores. Higher scores on the five index scores and the total scale score indicated better performance.

The CDR is a semi-structured interview used to stage the severity of dementia covering the following domains: memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care (Hughes et al., 1982). It was originally developed for use in people with Alzheimer's but can be used to stage dementia in other illnesses. A translation into Chinese has previously been reported with details of cultural modifications (Li, Ng, Kua & Ko, 2005). It is a five point scale where: 0 = no cognitive impairment; 0.5 = very mild dementia; 1 = mild dementia; 2 = moderate dementia; 3 = severe dementia. The CDR was

assumed to be 0 for those who had MMSE>=26, which was the study's cut-off for including the participants in further clinical assessments for dementia.

A panel comprised of 2 geriatricians, 1 neuropsychiatrist, 1 psychiatric epidemiologist and 3 clinical assessors reviewed the CDR classification, the clinical history, physical examination and laboratory investigations results, and the brain MRI scans to make a unanimous diagnosis of mild cognitive impairment (MCI) or dementia according to published criteria (Nyunt et al., 20130; Petersen et al., 1999). A diagnosis of cognitively normal was assumed for those with MMSE $\geq$ 26. This diagnosis is referred to as the clinical assessment of dementia (CAD) classification.

Statistical analysis

Participants were included in the analysis if they had completed all RBANS subtests and reported their age. Only RBANS assessments administered by a research nurse who conducted at least 10 RBANS interviews in both English and Chinese (Mandarin and/or dialects) were included. The latter inclusion criterion was intended to allow statistical control for prevention of interviewer effect that may confound the CDR (or CAD) comparison.

Demographic and baseline characteristics including gender, education level, language, CDR and CAD classification were summarized by counts and percentages; and continuous variables such as age, Geriatric Depression Scales (GDS), MMSE and physical activity scores were summarized by means and standard deviations (SD). The RBANS index scores and the total scale score were summarized by means and SDs across CDR classification and the CAD classification.

We have previously demonstrated that the different language-dialect versions of RBANS showed practically equivalent scores, with two exceptions: the attention index of the English

language version was not equivalent to any of the Chinese language versions (Mandarin, Hokkien, Teochew or Cantonese), and the visuospatial/constructional index of the English language version was not equivalent to the Hokkien or Teochew language versions (Phillips et al., 2015). Therefore, the RBANS attention index was summarized across each language and the visuospatial/constructional index was summarized by English, Mandarin and Cantonese combined and Hokkien and Teochew individually.

Distribution-based approaches. Distribution-based methods estimate the MCID based on some measure of variability of the observed scores (Copay et al., 2007). In this study we examined the following:

(1) MCID calculated as the standard error of measurement (SEM). This is a measure of the variation in scores due to the unreliability in the scale used. A smaller difference or change in score than the SEM is likely to be due to measurement error. It was calculated as the standard deviation (SD) of the measure multiplied by the square root of 1 minus the measure's reliability coefficient:

$$\text{SEM} = SD \times \sqrt{1 - r}$$

where r is the reliability as measured by Cronbach's alpha for this analysis. It has been noted that the value of 1 SEM corresponded to the MCID value when defined with the classic anchor-based method (Wyrwich, Nienaber, Tierney & Wolinsky, 1999; Wyrwich, Tierney & Wolinsky, 1999). Others have suggested different values. For example, Ware, Kosinski and Keller (1994) used 2 SEM. However, 1 SEM has become the generally accepted value to present as the MCID.

(2) MCID calculated using the standard deviation (SD) of the scores. SD is the variation among a group of scores. Norman, Sloan and Wyrwich (2003) found that the 0.5 SD corresponded to the MCID across a variety of studies.

Due to the non-equivalence of language versions previously discussed, we repeated each of the above distributional-based methods for each language individually for the attention index and the English, Mandarin and Cantonese combined and the Teochew and Hokkien individually for the visuospatial/constructional index.

Anchor-based approaches. We compared the RBANS index scores and total scale score of participants between the CDR no dementia group and the CDR very mild dementia group. Additionally, we compared the RBANS scores between the CAD cognitively normal diagnosis group and the CAD MCI diagnosis group. Regression models for each of the five RBANS index scores and the RBANS total scale score were fitted to compare scores between no dementia (CDR=0) and very mild dementia (CDR=0.5) to establish each indexes' (or total scale score) MCID. To remove potential confounding by differences in demographic characteristics (age, gender, and education level) covariates were included. Furthermore, to prevent confounding by interviewer effects, a mixed model approach was used to include the research nurse as a random-effect (Y. B. Cheung, 2014).

Due to the non-equivalence of language versions, we refitted the above model with an additional covariate for language version for the attention index and the visuospatial/constructional index. Language was included as a categorical variable with values presenting English, Mandarin, Hokkien, Teochew and Cantonese for the attention index; and values presenting English, Mandarin, Cantonese combined, and Hokkien and Teochew individually for the visuospatial/constructional index.

We repeated the above to compare the RBANS total scale score and index scores between those classed as cognitively normal and those as MCI according to the CAD classification.

All estimates of the MCID are given in the metrics of mean 100 and SD 15.

Results

Descriptive Summary

The sample consisted of 1,856 participants with complete RBANS data. The mean age of participants was 66.28 years (SD 7.54), they were predominantly female (62.77% compared to 37.23% males), with a fifth having received no formal education (18.48%) and just over two-thirds (70.96%) receiving 10 years or less. Cantonese was the most common language preference (26.24%), closely followed by Mandarin (25.38%), English (21.66%), Hokkien (20.26%) and Teochew (6.47%). Mean GDS was in the normal range (0.72 SD 1.48).

CDR classification was available for 1,802 (97.09%) participants, the majority of which were classified as having no dementia (86.26%), with 9.48% having very mild dementia. There were no cases of severe dementia in the sample. CAD classification was available for 1,745 (94.02%) participants, with the majority being classified as cognitively normal (89.33%), 3.56% were given a diagnosis of MCI and 1.13% were given a dementia diagnosis (Table 1).

Mean RBANS index scores and the total scale score decreased across all indexes as the CDR classification indicated increasing levels of cognitive impairment (Table 2). Mean RBANS index scores and the total scale score decreased across all indexes as the CAD classification moved from cognitively normal, to mildly cognitively impaired through to dementia (Table 3).

Minimum Clinically Important Difference

Distribution-based approach. The threshold according to the SEM approach for the total scale score was 5.26 and approximately 10 for each index; the exception was the threshold for the language index which was 14.84. The threshold estimated from the SD was approximately the same as the estimate from the SEM for the total scale score, whereas the estimates for the other indexes were at least one point smaller than the estimates from the SEM (Table 4).

Anchor-based approach. The MCID estimates with the CDR classification (no dementia vs. very mild dementia) were 7.79, 10.74, 8.63 and 9.74 for the total scale score, immediate memory index, language index and delayed memory index, respectively. The estimates for the visuospatial/constructional and attention indexes with language adjustment were somewhat smaller at 5.61 and 3.77, respectively (Table 5).

The MCID estimates with the CAD classification (cognitively normal compared to MCI) were 9.98, 12.16, 14.30 and 11.14, for the total scale score, immediate memory index, language index and delayed memory index, respectively. Again the estimates for the visuospatial/constructional and attention indexes were somewhat smaller at 6.88 and 4.58, respectively (Table 6). The estimates with the CAD classification were approximately no more than two points larger than the CDR estimates. The exception being the language index, which was over 5 points larger.

Overall. Figure 1 presents the estimates for the MCID for the immediate memory index, language index, delayed memory index and the total scale score according to each distributional and anchor-based approach. The anchor-based method using the CAD classification tended to give the largest estimate for the MCID and the distributional-based method using the SD gave the smallest estimate. The anchor-based method using the CDR classification and the SEM tended to be within 2 points of each other. Again, the exception to this was the language index where the SEM and anchor-based approach using the CAD gave

the largest scores and the anchor-based estimate using the CDR and the distributional estimate using the SD gave the smallest. Estimates for each index tended to span a 5 point width across estimates.

Figures 2 and 3 present the estimates for the MCID for the indexes that did not achieve language equivalence, the visuospatial/constructional index and the attention index. The SEM approach produced the largest estimate for each index. The anchor-based approach using the CDR produced the smallest with the other two estimates (SD and anchor-based approach with CAD) falling within two points of it.

Discussion

The RBANS is a useful battery to evaluate elderly patients with abnormal cognitive decline in both clinical trials and clinical practice. However, to date there is no objective clinical interpretation for the numerical scores, therefore understanding the MCID in this assessment is critical in assessing the effectiveness of a treatment or clinical significance/patient relevance. The aim of this study was to estimate the MCID of the RBANS for each index score and the total scale score, scored according to the widely accepted, modified criteria, suggested by Duff et al. (2003). We used both anchor-based and distribution-based analytical techniques to obtain the estimates, which could be used to compare patient groups, in for example, a randomized clinical trial setting. The study used data from a large cohort, the SLAS-II population based study of ageing, community dwelling adults in Singapore.

All estimates of the MCID are given in the metrics of mean 100 and SD 15. For the RBANS indexes that previously demonstrated equivalence in scores across language versions, immediate memory, language and delayed memory, the anchor-based method using the CDR gave estimates for the MCID of 10.74, 8.63 and 9.74, respectively. The estimate for the MCID for the total scale score was slightly smaller at 7.79. The thresholds estimated from the SEM

calculation yielded similar values. The exception to this was for the language index where the SEM estimate was more in line with the CAD anchor estimate. Not unexpectedly the estimates from the CAD anchor gave marginally larger thresholds. This was expected since the CDR anchor examined change between cognitively normal and very mild cognitive impairment, whereas the CAD minimum threshold was between cognitively normal and MCI, which by definition describes a greater level of cognitive impairment and a greater decrease in RBANS scores would be expected in this group.

The two exceptions to the equivalence in scores across language versions were the visuospatial/constructional index and the attention index. For these indexes we found that the anchor-based method using the CDR gave estimates of 5.61 and 3.77 respectively, with the CAD estimates only marginally larger but the SEM calculation yielded estimates of approximately 10 points.

The estimates from the distribution-based methods were similar for the total scale score (5.26 for the SEM method and 5.39 for the 0.5*SD) and the delayed memory index (9.34 for the SEM method and 8.13 for the 0.5*SD). Mathematically, when the reliability measure for the index is 0.75 (a moderate level of reliability), these two methods exactly agree. Cronbach's alpha, the measure of reliability used in this study, was 0.76 and 0.67 for total scale and delayed memory index, respectively. Hence, the similarity between the MCID estimates from the two methods. The largest differences in estimates were for the language index and attention index, which had smaller Cronbach's alpha values (0.31 and 0.21, respectively). These weaker reliability values are in line with other studies that have looked at the reliability of the RBANS indexes (McKay, Casey, Wertheimer & Fichenberg, 2007; Cheng et al., 2001).

Not unexpectedly, there was variation in the MCID estimates across indexes. For example, the estimates from the SEM method ranged from 5.26 to 14.84. The weaker reliability values for

the language and attention indexes would lead to larger differences being required to convince users that any change was not due to insufficient reliability, hence the larger MCID estimates for these indexes compared to the indexes with stronger reliability values. The MCID estimates produced by the anchor based estimates ranged from 3.77 to 10.74. These estimates were obtained by comparing two groups, those with normal cognition and those with very mild dementia, as defined by the CDR. In this community dwelling, Chinese sample we expected that Alzheimer's Disease would be the dominant subtype of pathology in the early stage dementia patients. Biologically there is more impairment in memory in Alzheimer's patients than other dementia types, hence greater MCID estimates were obtained for memory indexes than other indexes using this estimation method.

No study has previously examined the MCID for the RBANS. Patton et al. found a 3 point decrease in RBANS total score over a 1 year period in a group of cognitively intact, community dwelling, older adults (mean age 72.5 years) and Lee et al. found a 1.2 point increase and a 3.5 point decrease in RBANS total scale score over an 8 week period in elderly, (mean age 65.1 and 65.2 years, respectively), cognitively normal, English and Chinese speaking adults, respectively (Patton et al., 2005; Lee et al., 2013; Lee et al., 2015). This gives an indication of the changes one would expect to see in cognitively intact adults. Therefore, the larger estimates we obtained that distinguish between cognitively normal and very mild impairment are in keeping with these findings.

There were several limitations to this study. Firstly, there are many different methods to calculate the MCID. As we have shown, anchor-based methods produce different MCIDs depending on the external criterion used and the distribution-based methods depend on the measure of statistical variability used. Distribution-based methods also fail to account for the clinical importance/patient relevance, which is at the very core of MCIDs. However, Revicki et al. (2008) recommended that the MCID should be based on relevant anchor-based

approaches and this has subsequently been recommended by the US FDA in their guidance for industry on patient reported outcomes (Food and Drug Administration, 2009) and the IMMPACT statement (Dworkin et al., 2008). Therefore, our conclusions are in line with current guidelines.

Secondly, there was only cross-sectional data available for the analysis, so MCID estimates were based on between patient differences; estimates based on within patient differences were not possible. Therefore, these estimates would be more suited to discriminate between patient groups rather than assessing response to treatment (Kohn et al., 2014). For example, in a clinical trial setting where one is trying to determine if a statistically significant difference between treatment groups, is large enough to be considered clinically significant.

For the anchor-based estimates the use of mixed effects models, adjusting for differences in characteristics amongst participants and controlling for interviewer effects, would help reduce the risk of confounding i.e. a characteristic that may affect the RBANS scores other than the CDR (or CAD) classification. However, there may still be the risk of residual confounding (Y.B. Cheung, 2014). Also, the anchor-based approaches were limited to the study population that was assessed using CDR (97.09%) and diagnosed according to the CAD criteria (94.02%). Examination of the demographics found that those with missing CDR and CAD assessments were similar in terms of mean age and sex. However, a greater proportion of the non-completers had received no education (51.85% for the CDR and 52.73% for the CDR). The non-completers also had smaller mean MMSE scores for both the CDR and CAD (22.71 and 22.68, respectively). This raises the issue that our analysis is missing some of the participants with mild dementia. The distribution based estimates are concerned with the distribution of the RBANS scores, therefore by their very nature they do not account for any differences in characteristics amongst the participants. However, the distribution based estimates were similar

to the anchor based estimates based on the CDR, which suggests that in this population it is mild dementia not age or education or gender that is the key determinant of RBANS.

Recruitment was conducted via door-to-door visits and specifically targeted senior activity centers, elderly day care centers, and elderly homes across different community neighborhoods in Singapore. This was to ensure that the sample population was truly representative of the target population, i.e., Singaporean, elderly population. Our sample age ranged from 54 to 94 years of age, and the sex, education profile of participants closely followed the distribution of the Singaporean population aged over 55 years, according to the 2010 Singapore Census. However, one must also note that the use of a Singaporean population and use of Singaporean population norms for scoring the RBANS likely limits the generalizability, although Singapore normative data may have applicability to ethnic Chinese living outside of Asia.

In conclusion, based on current guidance we estimate the MCID to be 8 points for the RBANS total scale score, 9 points for the language index, 10 points for the RBANS immediate memory and delayed memory indexes and 6 and 4 points for the RBANS visuospatial/constructional and attention indexes, respectively. We recommend that the MCIDs estimated here are best suited to discriminate between patient groups, in for example, a clinical trial setting. At present, clinicians can use RBANS scores in clinical practice along with their judgment based on their clinical experiences. Further research is needed using longitudinal data to assess their applicability to assess within patient differences.

Acknowledgments

References

Badenes, G.D., Casas, H.L., Cejudo, B.J.C., & Aguilar, B.M. (2008). Evaluation of the capacity to drive in patients diagnosed of mild cognitive impairment and dementia. (Article in Spanish). *Neurologia*, 23(9), 575-582.

Boone, K.B., Victor, T.L., Wen, J., Razani, J., & Ponto´n, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, 22(3), 355-365. doi:10.1016/j.acn.2007.01.010.

Cheng, Y., Wu, W., Wang, J., Feng, W., Wu X., & Li, C. (2001). Reliability and validity of the Repeatable Battery for the Assessment of Neuropsychological Status in community dwelling elderly. *Archives of Medical Science*, 7(5), 850-857. doi:10.5114/aoms.2011.25561.

Cheung, Y.B., Thumboo, J., Goh, C., Khoo, K.S., Che, W., & Wee, J. (2004). The equivalence and difference between the English and Chinese versions of two major, cancer-specific, health-related quality-of-life questionnaires. *Cancer*, 101(12), 2874-2880. doi: 10.1002/cncr.20681.

Cheung, Y.B. (2014). Statistical analysis of Human Growth and Development. Chapman & Hall/CRC Biostatistics Series.

Cheung, Y.T., Foo, Y.L., Shwe, M., Tan, Y.P., Fan, G., Yong, W.S., … Chan, A. (2014). Minimal clinically important difference (MCID) for the functional assessment of cancer therarpy: Cognitive function (FACT-Cog) in breast cancer patients. *Journal of Clinical Epidemiology,* 67, 811-820. doi: 10.1016/j.jclinepi.2013.12.011.

Collinson, S.L., Fang, S.H., Lim, M.L., Feng, L., & Ng, T.P. (2014). Normative data for the Repeatable Battery for the Assessment of Neuropsychological Status in elderly Chinese. *Archives of Clinical Neuropsychology*, 29(5), 442-455. doi:10.1093/arclin/acu023.

Copay, A.G., Subach, B.R., Glassman, S.D., Polly, D.W., & Schuler, T.C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal*, 7, 541-546. doi:10.1016/j.spinee.2007.01.008.

Duff, K., Patton, D., Schoenberg, M.R., Mold, J., Scott, J.G., & Adams, R.L. (2003). Age and education-corrected independent normative data for the RBANS in a community dwelling elderly sample. *The Clinical Neuropsychologist*, 17(3), 351-366. doi:10.1076/clin.17.3.351.18082.

Duff, K., Schoenberg M.R., Patton, D., Paulsen, J.S., Bayless, J.D., Mold, J., … Adams, R.L. (2005). Regression-based formulas for predicting change in RBANS subtests with older adults. *Archives of Clinical Neuropsychology*, 20, 281, 290. doi:10.1016/j.acn.2004.07.007.

Dworkin, R.H., Turk, D.C., Wyrwich, K.W., Beaton, D., Cleeland, C.S., Farrar, J.T., … Zavisic, S. (2008). Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *The Journal of Pain*, 9(2), 105-121. doi:10.1016/j.jpain.2007.09.005.

Feng, L., Chong, M.S., Lim, W.S., & Ng, T.P. (2012) The Modified Mini-Mental State Examination test: normative data for Singapore Chinese older adults and its performance in detecting early cognitive impairment. *Singapore Medical Journal*, 53(7), 458-462.

Food and Drug Administration. (2009). Guidance for industry. Patient reported outcome measures: use in medical product development to support labelling claims. Federal Register 74(235), 65132-65133.

Fuji, D.E.M. (2010). The neuropsychology of Asian-Americans. London: Psychology Press.

Guyatt, G., Walter, S., & Norman, G. (1985). Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases*, 40(2), 171-178.

Guyatt, G.H., Osoba, D., Wu, A.W., Wyrwich, K.W., & Geoffrey, R. (2002). Clinical Significance Consensus Meeting Group, Methods to Explain the Clinical Significance of Health Status Measures. *Mayo Clinical Proceedings*, 77 (4), 371-383. http://dx.doi.org/10.4065/77.4.371.

Hedden, T., Park, D.C., Nisbett, R., Ji, L.J., Jing, Q. & Jiao, S. (2002). Cultural variation in verbal versus spatial neuropsychological function across life span. *Neuropsychology*, 16(1), 65-73. doi:10.1037//0894-4105.16.1.65.

Hermes, E.D.A., Sokoloff, D., Stroup, T., & Rosenheck, R.A. (2012). Minimum clinically important difference in the positive and negative syndrome scale with data from the clinical antipsychotic trials of intervention effectiveness (CATIE). *Journal of Clinical Psychiatry*, 73(4), 526-532. doi:10.4088/JCP.11m07162.

Hughes, C.P., Berg, L., Danziger, W.L. Coben, L.A., & Martin, R.L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140, 566-572. doi:10.1192/bjp.140.6.566.

Jaeschke, R., Singer, J., & Guyatt, G.H. (1989). Measurement of health status: ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, 10 (4), 407-415.

King, A.T. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics and Outcomes Research*, 11(2), 171-184. doi: 10.1586/erp.11.9.

Kohn, C.G., Sidovar, M.F., Kaur, K., Zhu, Y. & Coleman, C.I. (2014). Estimating a minimal clinically important difference for the EuroQoL 5-dimension health status index in persons with multiple sclerosis. *Health and Quality of Life Outcomes*, 12(66). doi:10.1186/1477-7525-12-66.

Kotani, S., Sakaguchi, E., Warashina, S., Matsukawa, N., Ishikura, Y., Kiso, Y., … Yamashima, T. (2006). Dietary supplementation of arachidonic and docosahexaenoic acids improves cognitive dysfunction. *Neuroscience Research*, 56, 159-164. doi:10.1016/j.neures.2006.06.010.

Lee, T.S., Goh, S.J.A., Quek, S.Y., Phillips, R., Guan, C., Cheung, Y.B., … Krishnan, K.R.R. (2013) A Brain-Computer Interface Based Cognitive Training System for Healthy Elderly: A Randomized Control Pilot Study for Usability and Preliminary Efficacy. *PLOS ONE*, 8(11), e79419. doi:10.1371/journal.pone.0079419.

Lee, T.S., Quek, S.Y., Goh, S.J.A., Phillips, R., Guan, C., Cheung, Y.B., … Krishnan, K.R.R. (2015). A pilot randomized controlled trial using EEG-based brain-computer interface training for a Chinese-speaking group of healthy elderly. *Clinical Interventions in Aging*, 10, 217-227. doi: 10.2147/CIA.S73955.

Li, M., Ng, T.P., Kua, E.H., & Ko, S.M. (2005). Brief informant screening test for Mild Cognitive Impairment and early Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, 21, 391-402. doi:10.1159/000092808.

Lim, M.L., Collinson, S.L., Feng, L., & Ng, T.P. (2010). Cross-cultural application of the Repeatable Battery for the assessment of Neuropsychological Status (RBANS): Performances of elderly Chinese Singaporeans. *The Clinical Neuropsychologist*, 24(5), 811-826. doi:10.1080/13854046.2010.490789.

Lydick, E., & Epstein, R.S. (1993). Interpretation of quality of life changes. *Quality of Life Research*, 2(3), 221-226.

Mahncke, H.W., Conner, B.B., Appelman, J., Ahsanuddin, O.N., Hardy, J.L., Wood, R.A., … Merzenich, M.M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences*, 103(33), 12523-12528. doi:10.1073_pnas.0605194103.

McKay, C., Casey, J.E., Wertheimer, J. & Fichenberg N.L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. *Archives of Clinical Neuropsychology*, 22, 91-98. doi:10.1016/j.acn.2006.11.003.

Ng, T.P., Niti, M., Chiam, P.C., & Kua, E.H. (2007). Ethnic and educational differences in cognitive test performance on Mini-Mental State Examination in Asians. *American Journal of Geriatric Psychiatry*, 15(2), 130-139. doi:10.1097/01.JGP.0000235710.17450.9a.

Norman, G.R., Sloan, J.A., & Wyrwich, K.W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical Care, 41(5), 582–92.

Nyunt, M.A.Z., Chong, M.S., Lim, W.S., Lee, T.S., Yap, P. & Ng, T.Z. (2013). Reliability and validity of the clinical dementia rating for community-living elderly subjects without an informant. *Dementia and Geriatric Cognitive Disorders Extra*, 31(1), 407-416. doi:10.1159/000355122.

Patton, D.E., Duff, K., Schoenberg, M.R., Mold, J., Scott, J.G. & Adams, R.L. (2005). Base rates of longitudinal RBANS discrepancies at one and two year intervals in community dwelling older adults. *The Clinical Neuropsychologist*, 19(1), 27-44. doi:10.1080/13854040490888477.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G. & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology*, 56(3), 303-308. doi:10.1001/archneur.56.3.303.

Phillips, R., Cheung, Y.B., Collinson, S.L., Lim, M.L., Ling, A., Feng, L. & Ng T.P. (2015). The equivalence and difference between the English and Chinese language versions of the Repeatable Battery for the Assessment of Neuropsychological Status. *The Clinical Neuropsychologist*. doi: 10.1080/13854046.2015.1034182.

Randolph, C. (1998). Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). San Antonio: Harcourt, TX: The Psychological Corporation.

Randolph, C., Tierney, M.C., Mohr, E., & Chase, T.N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity. *Journal of Clinical and Experimental Neuropsychology*, 20(3), 310-319. doi:10.1076/jcen.20.3.310.823.

Revicki, D., Hays, R.D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102-109. doi:10.1016/j.jclinepi.2007.03.012.

Shan, I.K., Chen, Y.S., Lee, Y.C., & Su, T.P. (2008). Adult normative data of the Wisconsin card sorting test in Taiwan. *Journal of the Chinese Medical Association*, 71(10), 517-522. doi:10.1016/S1726-4901(08)70160-6.

Ware, J.E., Kosinski, M., Keller, S.K. (1994). SF-36 physical and mental health summaries scales: a user's manual. Boston, MA: The Health Institute.

Wyrwich, K.W., Nienaber, N.A., Tierney, W.M., & Wolinsky, F.D. (1999). Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Medical Care*, 37(5), 469-478.

Wyrwich, K.W., Tierney, W.M., & Wolinsky, F.D. (1999b). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology*, 52 (9), 861–73. http://dx.doi.org/10.1016/S0895-4356(99)00071-2.

Table 1: Demographic characteristics (n=1856)

Variable		n	
Age – years, mean (SD)		1,856	66.28 (7.54)
GDS score, mean (SD)		1,856	0.72 (1.48)
MMSE, mean (SD)		1,842	28.11 (2.38)
Physical activity score, mean (SD)		1,791	8.93 (4.21)
Sex, n (%)			
	Male	691	(37.23)
	Female	1,165	(62.77)
Education, n (%)			
	None	343	(18.48)
	1-3 years	289	(15.57)
	4-6 years	524	(28.23)
	7-10 years	504	(27.16)
	More than 10 years	190	(10.24)
	Missing	6	(0.32)
Language, n (%)			
	English	402	(21.66)
	Mandarin	471	(25.38)
	Hokkien	376	(20.26)
	Teochew	120	(6.47)
	Cantonese	487	(26.24)
CDR, n (%)			
	0 No dementia	1,601	(86.26)
	0.5 very mild dementia	176	(9.48)
	1 mild dementia	23	(1.24)
	2 moderate dementia	2	(0.11)
	Missing	54	(2.91)
CAD, n (%)			
	Cognitively normal	1658	(89.33)
	MCI	66	(3.56)
	Dementia	21	(1.13)
	Missing	111	(5.98)

SD: standard deviation; GDS: geriatric depression scale; MMSE: mini-mental state examination; CDR: clinical dementia rating; CAD: clinical assessment of dementia; MCI: mild cognitive impairment.

Table 2: Repeatable Battery for the Assessment of Neuropsychological Status total scale score and index scores summarized by the Clinical Dementia Rating classification

	CDR classification									Overall		
	0 - No dementia			0.5 - Very mild dementia			1 - Mild dementia					
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Immediate Memory	1,601	97.84	13.13	176	85.20	14.40	25	73.43	15.85	1,802	96.26	14.08
Language	1,601	99.22	15.51	176	90.06	28.82	25	77.96	19.64	1,802	98.03	17.68
Delayed Memory	1,601	96.29	14.72	176	84.65	18.08	25	66.01	23.74	1,802	94.73	15.98
Total Scale	1,601	98.58	9.53	176	88.68	12.22	25	77.21	11.97	1,802	97.32	10.56
Visuospatial/Constructional	1,601	100.02	12.93	176	90.57	16.88	25	80.27	18.64	1,802	98.82	13.92
English/Mandarin/Cantonese	1,211	102.34	11.40	103	95.39	15.04	15	81.65	20.20	1,329	101.56	12.17
Hokkien	291	92.79	14.33	59	83.31	17.05	8	79.20	18.70	358	90.92	15.38
Teochew	99	92.98	15.70	14	85.72	17.82	2	74.18	8.29	115	91.77	16.13
Attention	1,601	99.55	11.91	176	92.91	11.85	25	88.36	12.54	1,802	98.75	12.13
English	370	99.20	13.26	25	88.80	11.74	1	79.72		396	98.49	13.41
Mandarin	439	101.57	10.98	23	95.84	13.00	2	88.22	10.10	464	101.23	11.16
Hokkien	291	96.51	10.91	59	89.71	12.02	8	83.46	11.06	358	95.10	11.49
Teochew	99	97.54	13.92	14	92.93	8.39	2	86.02	10.99	115	96.78	13.42
Cantonese	402	100.38	11.23	55	96.97	10.64	12	92.76	14.10	469	99.78	11.33

SD: standard deviation; CDR: clinical dementia rating scale.
NOTE: 54 participants did not have a CDR score.
NOTE: Mild dementia category includes two patients with moderate dementia.

Table 3: Repeatable Battery for the Assessment of Neuropsychological Status total scale score and index scores summarized by the Clinical Assessment of Dementia classification

	CAD classification									Overall		
	Cognitively normal			MCI			Dementia					
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Immediate Memory	1,658	97.45	13.39	66	82.75	13.72	21	72.47	12.76	1,745	96.60	13.93
Language	1,658	99.03	15.75	66	83.69	39.75	21	78.24	21.17	1,745	98.20	17.70
Delayed Memory	1,658	95.96	14.95	66	82.40	17.17	21	63.82	19.97	1,745	95.06	15.70
Total Scale	1,658	98.38	9.67	66	85.58	13.32	21	76.20	11.10	1,745	97.63	10.41
Visuospatial/Constructional	1,658	99.94	12.97	66	87.95	17.60	21	77.79	18.54	1,745	99.22	13.65
English/Mandarin/Cantonese	1,257	102.23	11.44	32	91.87	17.53	12	78.21	18.77	1,301	101.76	12.02
Hokkien	295	92.96	14.20	31	83.87	17.53	7	78.98	21.93	333	91.82	15.02
Teochew	106	92.12	16.27	3	88.21	13.92	2	71.08	3.90	111	91.63	16.27
Attention	1,658	99.51	11.93	66	91.09	11.33	21	88.68	14.02	1,745	99.06	12.09
English	382	98.91	13.28	8	90.27	12.76	1	79.72		391	98.69	13.33
Mandarin	450	101.63	10.96	6	88.73	14.19	2	88.22	10.10	458	101.40	11.10
Hokkien	295	96.48	11.18	31	88.37	10.66	7	86.89	17.74	333	95.52	11.57
Teochew	106	97.37	13.65	3	91.76	5.20	2	81.07	3.99	111	96.93	13.57
Cantonese	425	100.43	11.13	18	96.81	10.52	9	92.86	13.92	452	100.14	11.21

SD: standard deviation; CAD: clinical assessment of dementia; MCI: mild cognitive impairment.
NOTE: 111 participants did not have a CAD assessment.

Table 4: Distribution-based methods for estimates of the Minimum Clinically Important Difference for the Repeatable Battery for the Assessment of Neuropsychological Status total scale score and index scores

				Minimum Clinically Important Difference	
RBANS	N	Mean	SD	SEM	0.5*SD
Immediate Memory	1,856	95.96	14.19	9.25	7.09
Language	1,856	97.76	17.92	14.84	8.96
Delayed Memory	1,856	94.40	16.25	9.34	8.13
Total Scale	1,856	97.03	10.78	5.26	5.39
Visuospatial/Constructional	1,856	98.56	14.07	9.13	7.03
English/Mandarin/Cantonese	1,360	101.35	12.33	8.37	6.17
Hokkien	376	90.68	15.51	10.96	7.76
Teochew	120	91.63	15.99	10.80	7.99
Attention	1,856	98.47	12.30	10.90	6.15
English	402	98.37	13.45	10.34	6.73
Mandarin	471	101.10	11.28	9.29	5.64
Hokkien	376	94.94	11.64	10.00	5.82
Teochew	120	96.24	13.59	9.35	6.80
Cantonese	487	99.29	11.71	9.17	5.85

SEM: standard error of measurement; SD: standard deviation.

Table 5: Regression analysis for estimates of the Minimum Clinically Important Difference to discriminate between a Clinical Dementia Rating classification of no dementia and a Clinical Dementia Rating classification of very mild dementia for the Repeatable Battery for the Assessment of Neuropsychological Status total scale score and index scores (n=1769)

RBANS	Mean	SD	Minimum Clinically Important Difference		
			Adjusted Difference [a]	95% CI	
Immediate Memory	96.25	14.09	10.74	8.74	12.75
Language	98.00	17.69	8.63	5.89	11.38
Delayed Memory	94.73	15.97	9.74	7.37	12.12
Total Scale	97.32	10.56	7.79	6.32	9.26
Visuospatial/Constructional [b]	98.85	13.90	5.61	3.69	7.54
Attention [b]	98.74	12.13	3.77	2.05	5.49

CDR: clinical dementia rating scale; SD: standard deviation; CI: confidence interval.
[a] Difference between CDR no dementia vs. CDR very mild dementia, adjusted for age, gender and education level. Research nurse included as a random effect. Differences indicate a decrease in score (worsening scores) from CDR no dementia to CDR very mild dementia.
[b] Additionally adjusted for language.
NOTE: 54 participants did not have a CDR score and a further 6 participants did not have any education data.

Table 6: Regression analysis for estimates of the Minimum Clinically Important Difference to discriminate between a Clinical Assessment of Dementia classification of cognitively normal and a Clinical Assessment of Dementia classification of mild cognitive impairment for the Repeatable Battery for the Assessment of Neuropsychological Status total scale score and index scores (n=1740)

RBANS	Mean	SD	Minimum Clinically Important Difference		
			Adjusted Difference [a]	95% CI	
Immediate Memory	96.59	13.94	12.16	9.02	15.30
Language	98.17	17.71	14.30	10.04	18.56
Delayed Memory	95.06	15.69	11.14	7.48	14.80
Total Scale	97.62	10.41	9.98	7.70	12.25
Visuospatial/Constructional [b]	99.22	13.66	6.88	3.93	9.84
Attention [b]	99.05	12.08	4.58	1.90	7.26

CAD: clinical assessment of dementia; MCI: mild cognitive impairment; SD: standard deviation; CI: confidence interval.

[a] Difference between CAD cognitively normal vs. CAD MCI, adjusted for age, gender and education level. Research nurse included as a random effect. Differences indicate a decrease in score (worsening scores) from CAD cognitively normal to CAD MCI.

[b] Additionally adjusted for language.

NOTE: 111 participants did not have CAD assessment and an additional 5 participants did not have any education data.

Figure 1: Range of Minimum Clinically Important Differences for the Repeatable Battery for the Assessment of Neuropsychological Status index scores and total scale scores from distributional and anchor-based estimations for equivalent language indexes.

Note: MCID: minimum clinically important difference; CDR: clinical dementia rating scale; CAD: clinical assessment of dementia; SEM: standard error of measurement; SD: standard deviation.

Figure 2: Range of Minimum Clinically Important Differences for the Repeatable Battery for the Assessment of Neuropsychological Status index scores and total scale scores from distributional and anchor-based estimations for visuospatial/constructional index.

Note: MCID: minimum clinically important difference; CDR: clinical dementia rating scale; CAD: clinical assessment of dementia; SEM: standard error of measurement; SD: standard deviation.

Figure 3: Range of Minimum Clinically Important Differences for the Repeatable Battery for the Assessment of Neuropsychological Status index scores and total scale scores from distributional and anchor-based estimations for attention index.

Note: MCID: minimum clinically important difference; CDR: clinical dementia rating scale; CAD: clinical assessment of dementia; SEM: standard error of measurement; SD: standard deviation.